

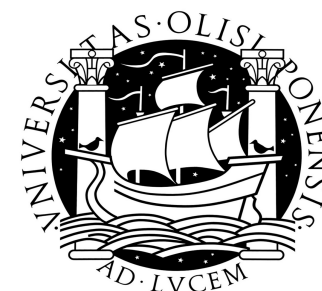


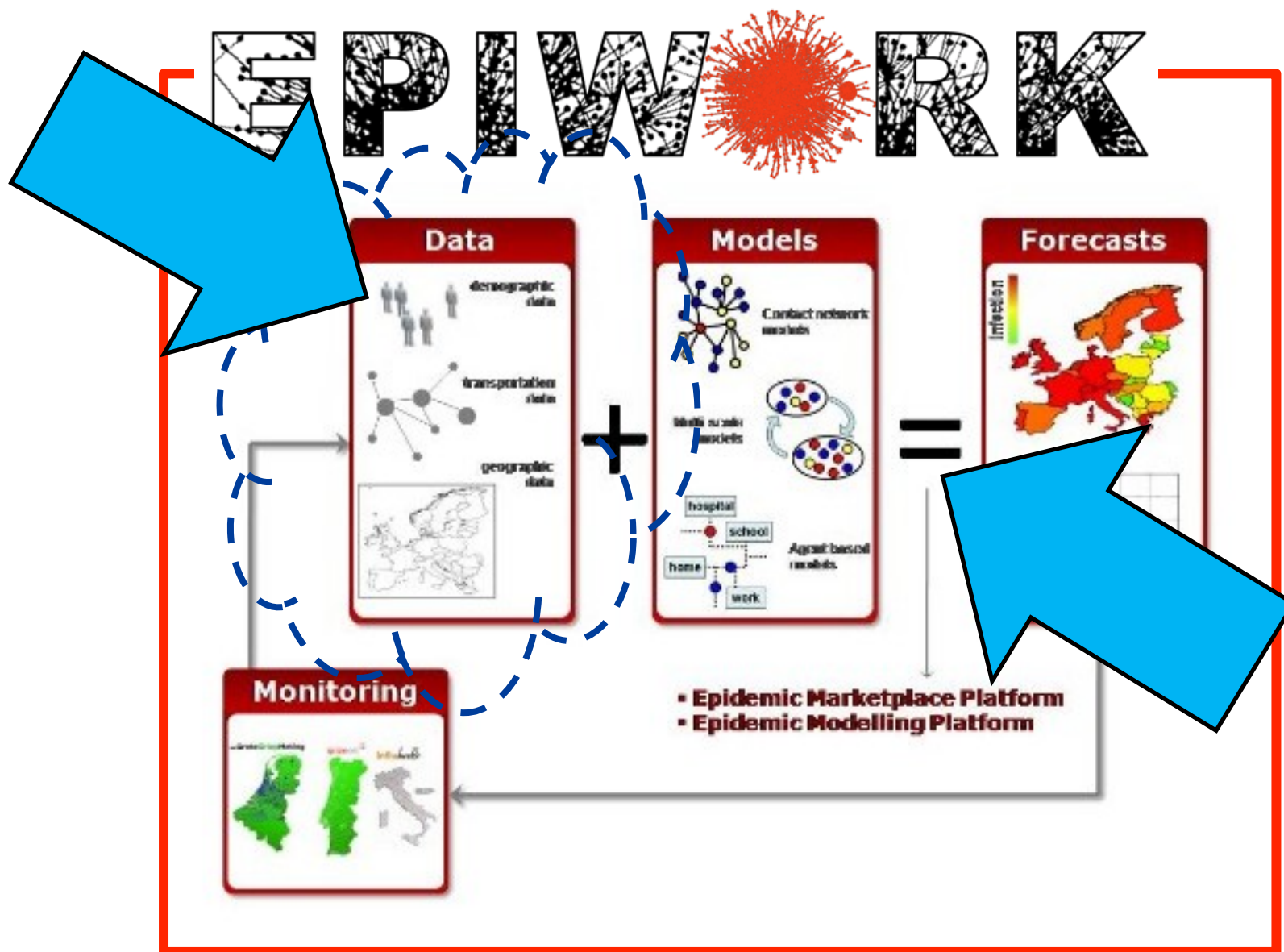
**WP3 – Information Platform  
Progress Report  
2<sup>nd</sup> Epiwork Meeting  
Torino 6<sup>th</sup>&7<sup>th</sup> Dec, 2010**

Mário J. Silva

Universidade de Lisboa, Faculdade de Ciências  
Departamento de Informática

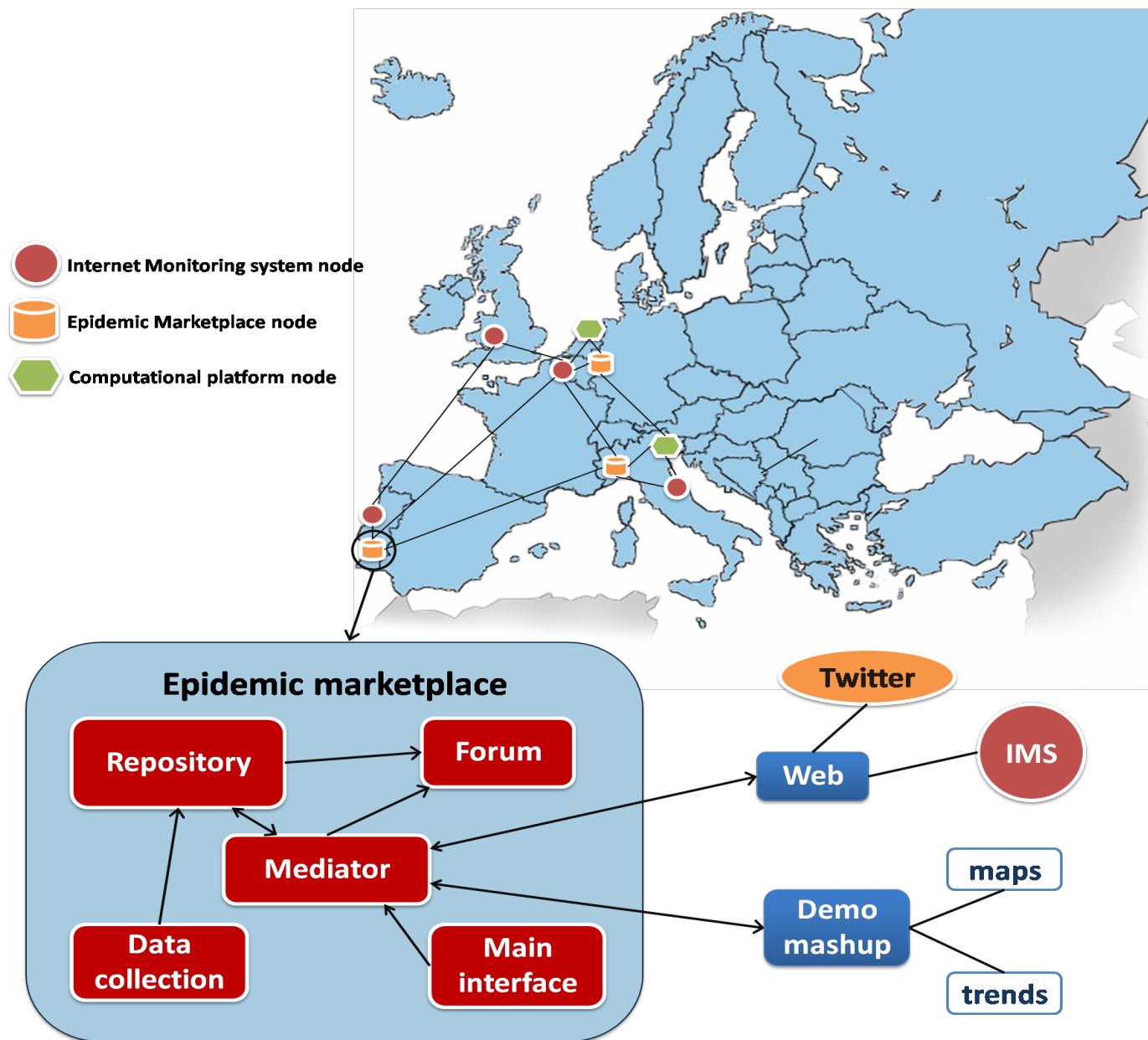
[mjs@di.fc.ul.pt](mailto:mjs@di.fc.ul.pt)





# Epidemic Marketplace (EM)

1. **Catalogue of data sources** containing the metadata describing existing databases;
2. **Forum** to
  - publish information about data
  - seek modellers to collaborate with,
  - seek sources of data that could be of interest to their epidemiological modelling efforts;
3. **Mediating software** to automatically process queries to epidemiological data, harvest data, assemble datasets....



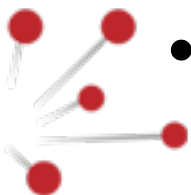
# Changes in UL WP3 Team

## Out

- Fabrício Silva
- Luis F. Lopes (meta-data)
- Hugo Ferreira (mediator)

## In

- Dulce Domingos (access control)
  - Juliana Duque (information architecture, graphics)
  - João Ferreira (ontologies)
- + (always in)
- Mário
  - Francisco
  - João Zamite



# EM 1.0 Software Components

- Fedora Commons 2.X for the implementation of the main features of the repository.
- Access control in the platform
  - XACML (OASIS 2010),
  - LDAP (Tuttle et al. 2004)
  - Shibboleth (identity management).
- Front-end based in Muradora
- Forum based on phpBB



[Home](#)[Browse](#)[Search](#)[Submit](#)[Publish](#)[Portfolio](#)[Admin Tools](#)[Home](#)

## Resource Metadata

Meta-data	
<b>Title</b>	Twitter dataset H1N1 + Portugal 4-6-2009
<b>Creator</b>	<a href="#">LASIGE node of the Epidemic Marketplace</a>
<b>Subject</b>	<a href="#">twitter message dataset</a>
<b>Description</b>	<p>This dataset contains Twitter messages containing the words H1N1 and Portugal collected between 16-5-2009 and 3-6-2009, Information is a 7 columns relation, containing the following data:</p> <p>Column 1- keyword 1 (disease)- H1N1            Column 2- Keyword 2 (location)- Portugal            Column 3- Source (Twitter)            Column 4- Author of the message (user id)            Column 5- The message body (evidence)            Column 6- score            Column 7- date (day and hour)</p>
<b>Publisher</b>	Epiwork – <a href="http://www.epiwork.eu">http://www.epiwork.eu</a>
<b>Format</b>	text/tab-separated-values
<b>Language</b>	English, Portuguese
<b>Contributor</b>	Luis F Lopes, Joao M Zamite, Bruno C Tavares, Francisco M Couto, Fabricio Silva, Mario J Silva
<b>Relation</b>	Luis F. Lopes, João M. Zamite, Bruno C. Tavares, Francisco M. Couto, Fabrício Silva and Mário J. Silva. (2009). Automated Social Network Epidemic Data Collector. INForum informatics symposium.
<b>Source</b>	<a href="http://epiwork.di.fc.ul.pt/collector/">http://epiwork.di.fc.ul.pt/collector/</a>
<b>Coverage</b>	Spatial: Portugal, Temporal: 2009-5-16 to 2009-6-3
<b>Rights</b>	Creative Commons Attribution-ShareAlike (CC BY-SA), <a href="http://creativecommons.org/licenses/by-sa/3.0/">http://creativecommons.org/licenses/by-sa/3.0/</a>

Old Graphic Style



# Epidemic Marketplace

// Epidemic Marketplace

// Datasets

// Forum

// Web Services

// Data Collection

// Contact

// Learn More

**Epidemic  
Marketplace**

*... a platform for integrating and sharing epidemiological data.*

Try it at:  
<http://epimarketplace.net>

Some features of the Epidemic Marketplace require user authentication. Please [register](#)

**Datasets**

**Forum**

**Web Services**

**Data Collection**

## Announcements

2010.10.01 // Epimarketplace.net is now open to user registrations from the entire community. [\(See More\)](#)

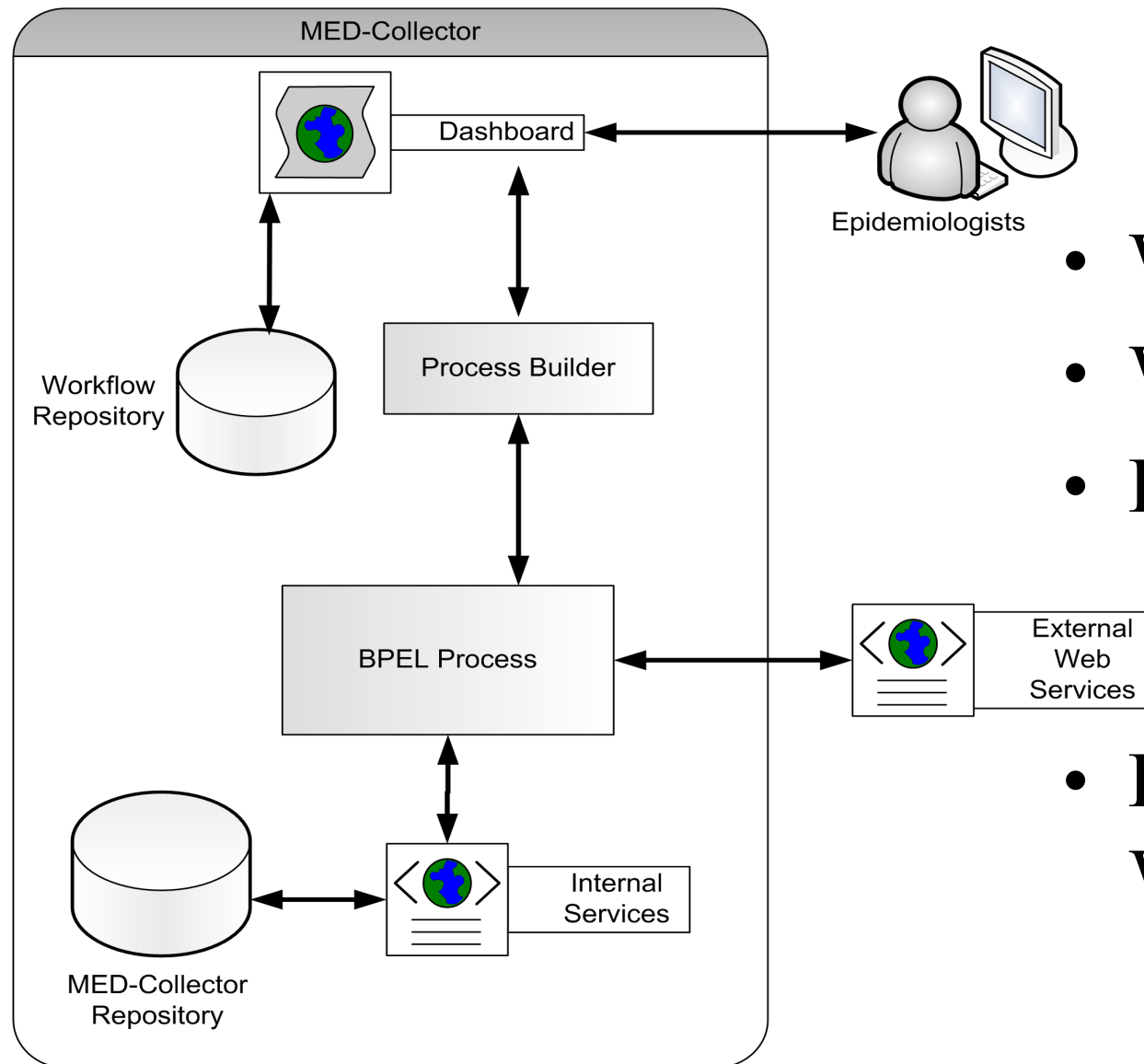


# **Deployment**

## **(EM Open to the Public D3.3)**

- Virtualized environment: every major component running on two separate virtual machines - production + development environments (Xen+CentOS)
- Monitoring and alerts for all services (Nagios)
- Logging and Analysis (Google Analytics)

# MEDCollector Architecture

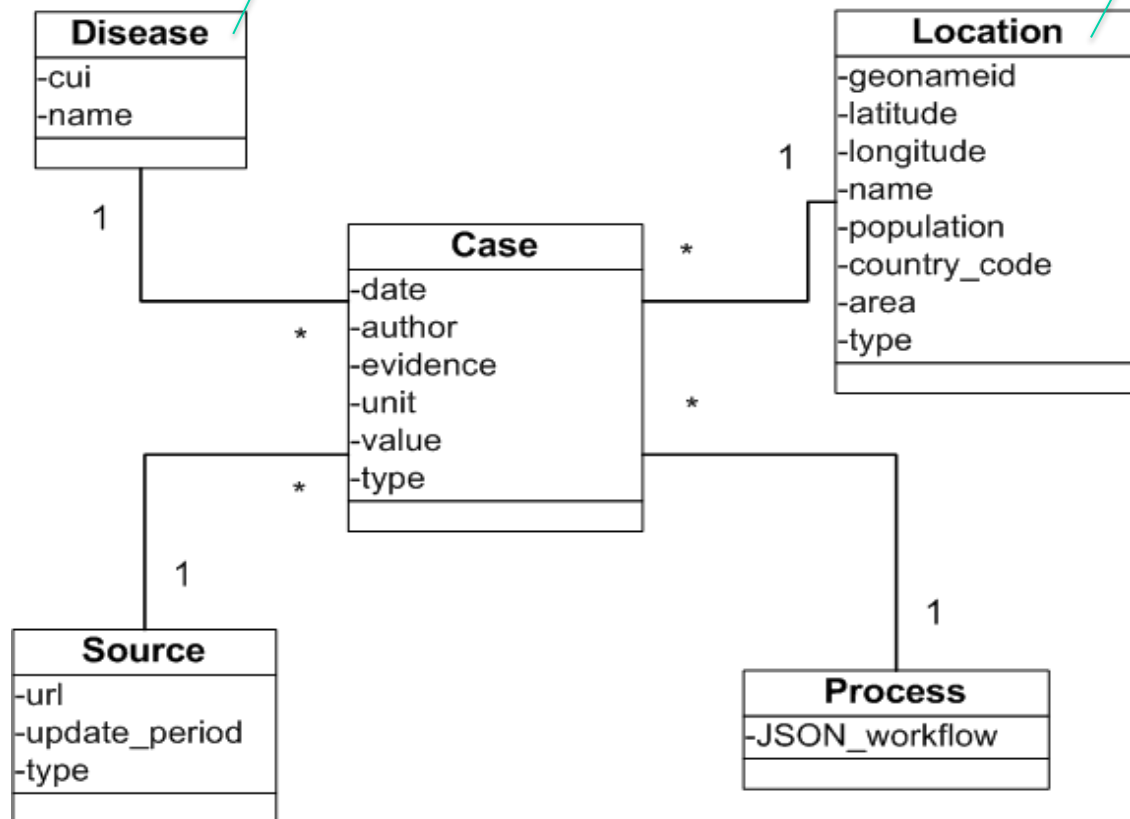


- Web Services
- Workflow Processes
- Local Storage
  - MEDCollector Repository
- Dashboard for Workflow Design

# MEDCollector Data Model

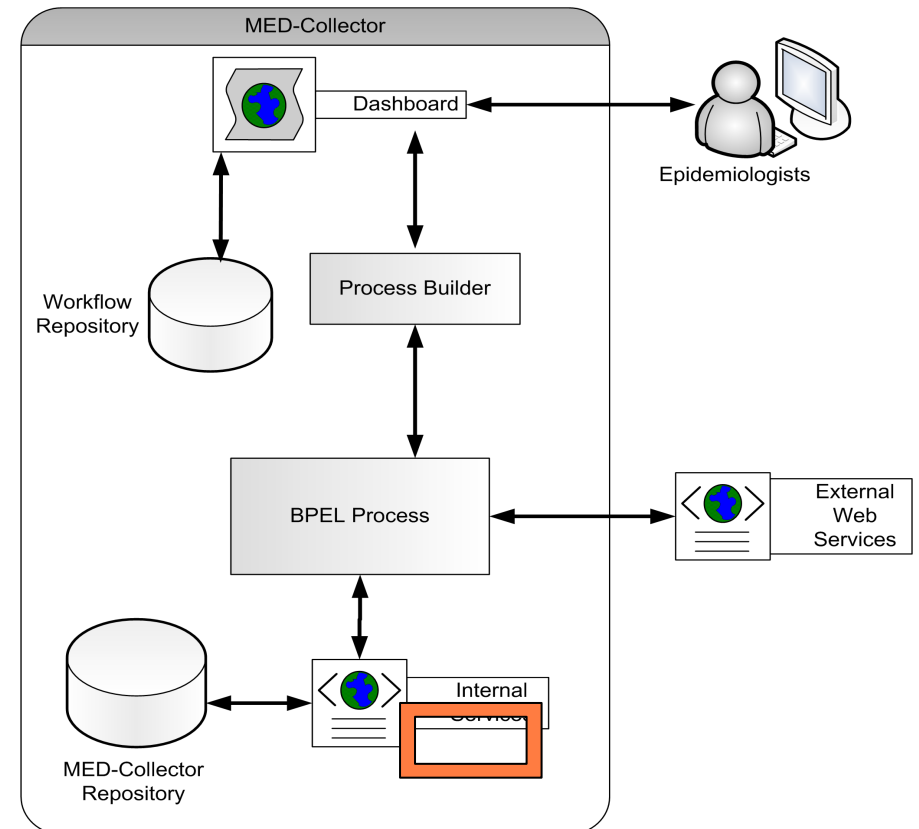
**UMLS:** twitter searched  
ion subset of 89 terms  
related to “Disease or  
Syndrome”

**Geonames.org:**  
All Countries  
and Capitals



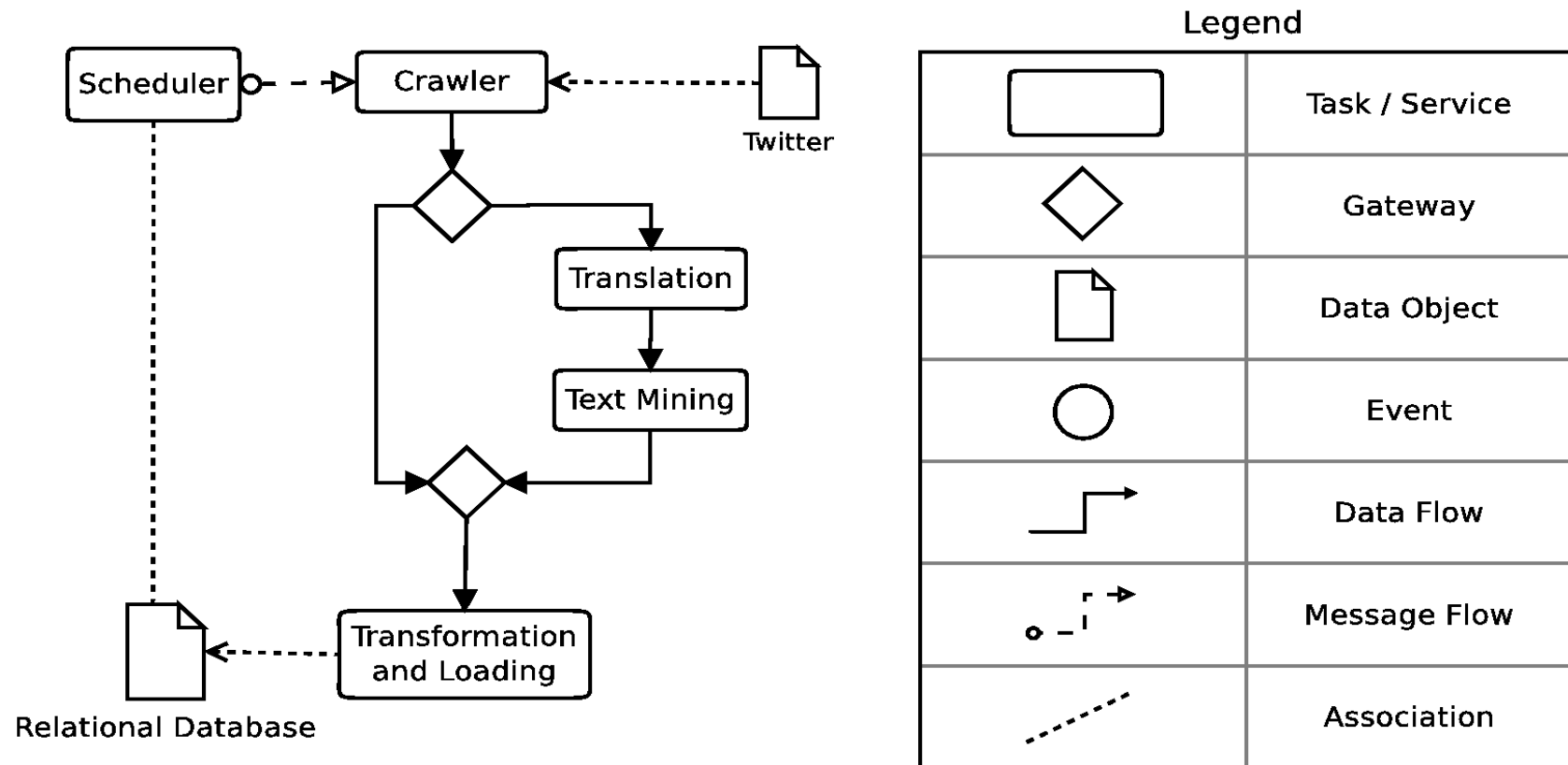
# MEDCollector Services

- Data Collection Services
  - The Query Selection Services
  - Data Harvesting Services
  - XML Transformation Services
  - Database Loading Service
- Data Packaging Services
  - To CSV



# MEDCollector - BPEL

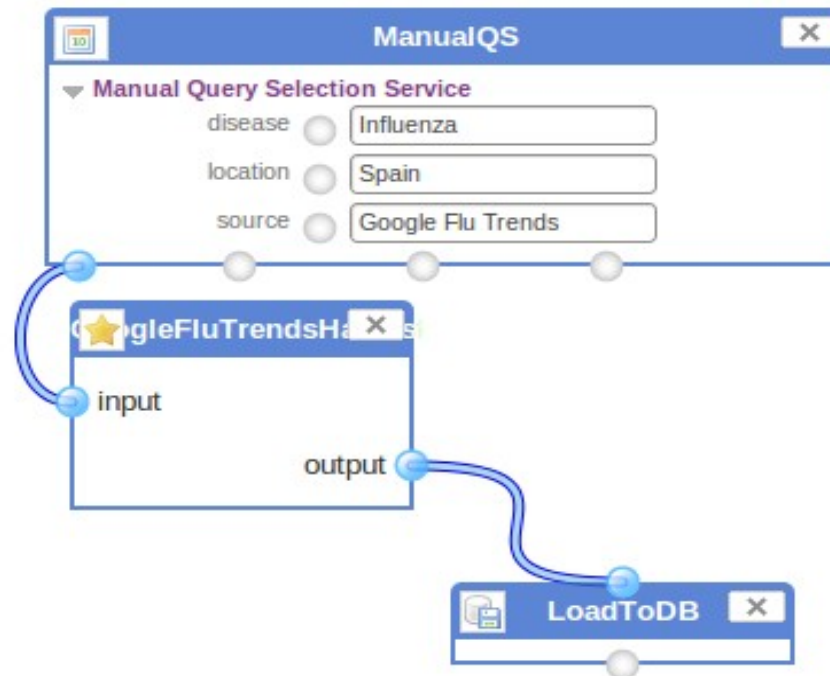
Language to define how Web-Services Communicate



- **Standard graphical notation – BPMN → Complex!**

# MEDCollector: Dashboard

WiteIt! - <http://javascript.neyric.com/wireit/>



# MEDCollector: Dashboard

The screenshot displays the MEDCollector Dashboard interface. At the top, the browser address bar shows `epiwork.dmr.cdc.gov/medcollector/index.php`. The main header includes navigation tabs: **Data Collection** (active), **Data Packaging**, and **View Workflows**. The page title is **Epidemic Marketplace Data Collection**, and a welcome message for João Zamite is visible along with [logout](#) and [Epidemic Marketplace](#) links.

The **Data Collection** section features a toolbar with icons for **New**, **Load**, **Save**, **Run**, **Delete**, **Help**, and **Feedback**.

On the left, a **Modules** sidebar lists options: **Query Selection**, **Crawlers**, **XML**, **Text**, **Loading**, **Date**, and **Other**.

The central workspace contains two modules:

- RESTfulWebService**: Includes a section for "Insert a URL and other parameters (for hybrid services)". The **URL** field is populated with `http://www2c.cdc.gov/podcasts/createrss.asp?t=r&c=20`. Below are fields for **Query**, **Parameter Name 1**, **Parameter Value 1**, **Parameter Name 2**, **Parameter Value 2**, **Parameter Name 3**, **Parameter Value 3**, **Parameter Name 4**, and **Parameter Value 4**.
- RSSToRecord**: A sub-module with the instruction "Disease and Location are optional." It contains fields for **Disease** (Influenza), **Location** (United States), and **Source** (CDC Flu updates).

On the right, the **Properties** panel shows configuration for **CDCFluRSS**. It includes a **Description** field, a **Scheduling** dropdown set to **Automatic**, and a **period (minutes)** field set to **720**. Below this is a **Minimap** section and a **View/Add Entities** table:

View/Add Entities	
<a href="#">View Diseases</a>	<a href="#">Add Disease</a>
<a href="#">View Sources</a>	<a href="#">Add Source</a>

**Watch the Demo!**



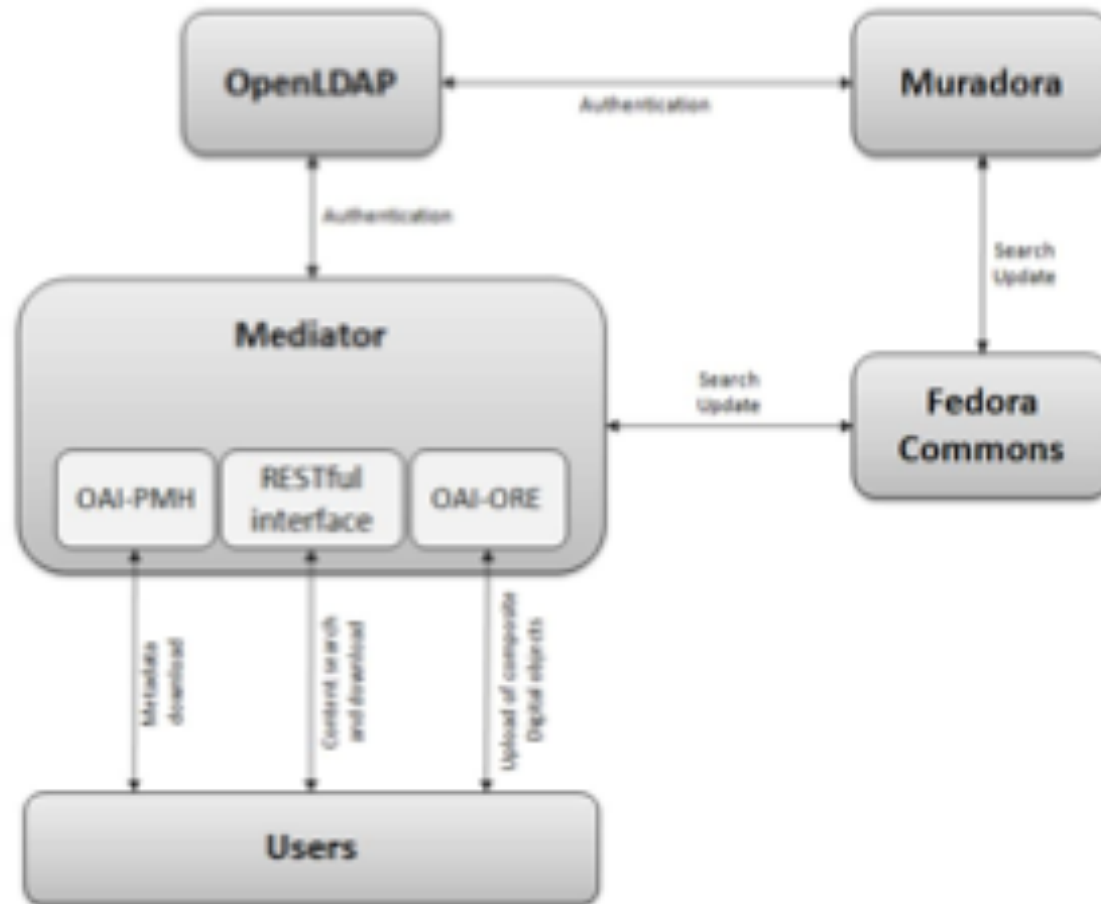
# Automatically Collected Data

- Twitter: 89 diseases, all over the world
- ProMed-mail
- Google Flu Trends
- CDC RSS Feeds
  - Flu updates
  - Travel Notices
  - ...
- Periodically packed and uploaded to the EM repository

# Mediator Web Services

7

- Upload
- Fetch
- Search



**Coming Soon...**

# Simple EM Client

- Mapping of client filenames to EM resources (FC data streams)
- Operations: Check-out, check-in, synchronization
- Later-on: mapping of directories to collections and dataset synchronization

# Meta-Data

- Meta-data introduction **simplified**
  - editor invoked at upload now fills most of the entries with defaults.
- EM Repository Meta-data Vocabulary
  - Generic DCTERMS for datasets characterisation
  - Epidemics-specific DCTERMS **defined** for epidemic datasets characterisation

# DC Term Example

<i>Name of Term</i>	<i>RightsHolder</i>
EM Term URI	<a href="http://epiwork.di.fc.ul.pt/em-terms/rightsHolder">http://epiwork.di.fc.ul.pt/em-terms/rightsHolder</a>
EM qualified name	em:rightsHolder
DC Term URI	<a href="http://purl.org/dc/terms/rightsHolder">http://purl.org/dc/terms/rightsHolder</a>
DC Qualified name	dc:rightsHolder
Label	Rights Holder
Defined By	<a href="http://dublincore.org/documents/dcmi-terms">http://dublincore.org/documents/dcmi-terms</a>
DC Definition	A person or organization owning or managing rights over the resource.
EM Definition	
DC Comments	
EM Comments	
Refines	<a href="http://epiwork.di.fc.ul.pt/em-terms/rights">http://epiwork.di.fc.ul.pt/em-terms/rights</a>
Refined By	
Has Encoding Scheme	
Obligation	<b>RA</b>
Occurence	0 or more

# EM Term Example

<i>Name of Term</i>	<i>HostGroup</i>
Term URI	<a href="http://epiwork.di.fc.ul.pt/em-terms/hostGroup">http://epiwork.di.fc.ul.pt/em-terms/hostGroup</a>
Qualified name	em:hostGroup
Label	Host Group
Defined By	<a href="http://epiwork.di.fc.ul.pt/em-terms">http://epiwork.di.fc.ul.pt/em-terms</a>
EM Definition	A dataset may contain data not about a specific species but rather a larger group comprising different species. When the larger group is the target this element should be used. For example, when referring to groups such as rodents or household animals.
EM Comments	When possible a scientific classification should be used.
Refines	<a href="http://epiwork.di.fc.ul.pt/em-terms/epidemiological">http://epiwork.di.fc.ul.pt/em-terms/epidemiological</a>
Refined By	
Has Encoding	
Scheme	
Obligation	<b>RA</b>
Occurence	0 or more

# Events 2nd year

London, Delhi, Bilbao, ERCIM News

**ECCS'10 LISBON**  
European Conference on Complex Systems  
eccs2010.eu



## Privacy in Socially Intelligent

Mário J. Silva

Universidade de Lisboa, Faculdade de Ciências,  
Departamento de Informática

[mjs@di.fc.ul.pt](mailto:mjs@di.fc.ul.pt)



13-17 September 2010

Lisbon University Institute



Towards a Science of Socially Intelligent ICT  
Imperial College London, 3rd August 2010



# EM-related Publications (2nd year)

1. Mário J. Silva, Fabrício A.B. Silva, Luís Filipe Lopes, Francisco M Couto, Building a Digital Library for Epidemic Modelling. Proceedings of ICDL 2010 - The International Conference on Digital Libraries 1, p. 447–459, New Delhi, India, 23–27 February, 2010. TERI Press–New Delhi, India. Invited Paper.
2. Fabrício A.B. Silva, Mário J. Silva, Francisco M Couto, Epidemic Marketplace: an e-Science Platform for Epidemic Modelling and Analysis. ERCIM News 82 – Special Theme: Computational Biology. July, 2010.
3. Luis Filipe Lopes, Fabrício A.B. Silva, Francisco M Couto, João Zamite, Hugo Ferreira, Carla Sousa, Mário J. Silva, Epidemic Marketplace: An Information Management System for Epidemiological Data. Proceedings of ITBAM'10 - 1st International Conference on Information Technology in Bio- and Medical Informatics - DEXA 2010 August, 2010.
4. João Zamite, Fabrício A.B. Silva, Francisco M Couto, Mário J. Silva, MEDCollector: Multisource Epidemic Data Collector. Proceedings of ITBAM'10 - 1st International Conference on Information Technology in Bio- and Medical Informatics - DEXA 2010 August, 2010.
5. João Zamite, Multisource Epidemic Data Collector, Master Dissertation, University of Lisbon, Faculty of Sciences, September 2010.
6. Luis Filipe Lopes, A Metadata Model for the Annotation of Epidemiological Data, Master Dissertation, University of Lisbon, Faculty of Sciences, September 2010.
7. Hugo Ferreira, O Mediador do Epidemic Marketplace. Master Dissertation, University of Lisbon, Faculty of Sciences, September, 2010; (in Portuguese).

# WP3: status

- D3.1 (meta-model) released
  - D3.2 (prototype) released
  - D3.3 (public) released
  - D3.4 and D3.5 due Feb 2012
- 
- Overcoming the initial difficulties in hiring the planned resources.
  - Refreshed team with competencies required for the 2nd and 3rd year.

# Current Focus

- Refining and populating, *enriching the catalogue* of epidemic resources
- Demonstrate integration with Computational Platform - **Simple EMClient**
- Designing and implementing new **social network access control**
- Designing the the second version of the EM.
  - New information architecture, new front-end design
  - Must be useful to the expert and occasional user.

# EM 2.0 Software Components

- Fedora Commons 3.X - main features of the repository.
  - Mediator services reimplemented calling webservices provided by FC
- Access control in the platform
  - XACML + LDAP (Tuttle et al. 2004)
  - Shibboleth (identity management).
- Front-end based in the **Drupal** CMS
  - Integrated forum



# Current Challenges

- Motivate the community to populate the Epidemic Marketplace.
  - Chicken and egg situation.
- **Data anonymization** is a major concern
  - Rights management to the sentence level!
  - Anyone giving away curated UGC?
- Dataset selection and generation policies

**Unchanged !**

# Forthcoming Developments

- Identifying ontologies (and ontology terms) to use. Linking to ontology definition initiatives.
- Linking ontologies and web data **using linked data** conventions and **ontology alignment** methods.



# **WP3 - Wrap up discussion and timelines before the project review meeting**

2nd Epiwork Meeting

Torino Dec 7th 2010



# WP3 SWOT Analysis

## Strengths

- Epiwork-driven EM
- Standards-based
- Open Source modules
- Supported (until 2012)

## Weaknesses

- **Unpopulated EM**
- Looking for the right policies
- What are the incentives?
- Interfaces to WP4 and WP5?

**Unchanged !**

# WP3 SWOT Analysis

## Opportunities

- Epiwork testbed
- Creation of a baseline for epidemic modelling
- Showcase for partners' outputs

## Threats

- Consortium enters “everyone for himself” mode.
- “Somebody will take care of that” attitude
- EM perceived as a very expensive, complex and useless cache

**Unchanged !**

# ToDo until Feb 2011

- Demo check-in/check-out integration with GleanViz
- 1st implementation of Social access control
- New user interface (EM 2.0) available internally

# **Todo list and planning (Torino, Dec 2010)**

1. Populate Repository
2. New Front-end and updated components  
(prepare market)
3. Access control policies & distributed authentication
4. Linked Epidemic Data
5. Ethics, Privacy and Anonimization
6. Replicate EM node

# Scheduled Deliverables

		Year 1			Year 2			Year 3			Year 4		
		M4	M8	M12	M16	M20	M24	M28	M32	M36	M40	M44	M48
<b>WP3</b>	<b>Information platform</b>												
Task 1			<b>D3.1</b>										
Task 2			<b>D3.1</b>										
Task 3				<b>D3.2</b>		<b>D3.3</b>				<b>D3.4</b>			<b>D3.6</b>
Task 4										<b>D3.4</b> <b>D3.5</b>			<b>D3.6</b>

7 Dec 2010